Content Filtering Technologies and Internet Service Providers

Enabling User Choice

Michael Shepherd and Carolyn Watters[1]

Web Information Filtering Lab
Faculty of Computer Science
Dalhousie University
6050 University Avenue
Halifax, Nova Scotia, Canada B3H 1W5
{shepherd | watters}@cs.dal.ca

March 22, 2000

[1] With Margo Boyd, Research Assistant

Table of Contents

Executive Summary 3
1. Introduction 6
2. Classification Scheme for Filtering Technologies 9
2.1. Two Layer Scheme 9
2.1.1. Core Technologies 10
2.1.2. Filtering Applications 10
2.2. Access Modes 13
3. Description of Filtering Technologies 15
3.1 Core Technologies 15
3.1.1 Site Labeling Systems 15
3.1.2 URL lists 23
3.1.3 Automated Text Analysis 25
3.1.4 Authorization Control Techniques 31
3.1.5 Activity Tracing 31
3.2 Application Level 33
3.2.1 Browser Applications for Children 33
3.2.2 Child-Friendly Search Engines and Portals 33
3.2.3 Proxy Applications 34
3.2.4 Activity Monitoring Applications 36
3.2.5 Restricted Access Applications 37
3.2.6 Non-HTTP Applications 37
4. Assessment of Types of Filtering Technologies 39
4.1 Core Technologies 39
4.1.1 Site Labels 39
4.1.2 URL lists 40
4.1.3 Automated Text Analysis 40
4.1.4 Authorization Techniques 41
4.1.5. Activity Tracing 42
4.2 Application technologies 43
4.2.1 Special Purpose Browsers for Children 43
4.2.2. Child Friendly Search Engines and Portals 43
4.2.3. Proxy Applications 43
4.2.4. Access Authority Applications 44
4.2.5. Activity Monitors 44
4.2.6. Filtering for non-HTTP Applications 44
5. Where the Filtering is done 46
5.1 Stand Alone Mechanisms (client machine) 46
5.2 ISP Mechanisms 47
6. Summary and Recommendations 49
Appendix 51
Examples of Filtering Software 51

Executive Summary

The project reports on the mechanisms that Internet Service Providers (ISPs) have the *option* to provide and that users can *choose* to utilize in order to filter the content delivered to users over the Internet and to allow authorized access to that content. The report is purely descriptive of the filtering mechanisms available and does not provide policy or legal advice or recommendations. It was commissioned by Industry Canada to help promote the development, awareness and use of tools and technologies that enable Internet users to make choices about the content that they access on the Internet.

Classification of Mechanisms

The mechanisms are classified into two tiers, the application level mechanisms and the underlying core technologies. The core technologies are classified as follows:

- **Site labels.** Labeling refers to schemes to assign content related labels to URL's and/or specific Web pages. The URL (Universal Resource Locator) describes the location of a specific Web page. Individual rating protocols exist, in general, separate from products or applications using these ratings. In general, these labels can be stored as part of the Web page or separately from the Web page in a database. Labels may be the result of self-rating, third-party authority rating, or community rating by interested users[2].
- **Lists of appropriate or inappropriate sites ("white" and "black" lists).** The most frequently used filtering technology is the use of lists of acceptable and/or unacceptable URL's. "White" lists are used to define a domain of "safe" Web sites within which users can browse. These typically require people to search and select sites that are approved by the provider of the list. "Black" lists are lists compiled of URL's from which requests will not be serviced. The lists are compiled as a service by individuals or by communities of raters.
- **Automated text analysis**. Another way to analyze a Web site is to use software that scans the text of a site to determine the relevance or suitability of pages. Users or groups of users have profiles of interests (positive and/or negative), consisting of keywords and phrases, that are used in this determination. Almost all content based filtering uses some variation of keyword matching, where keywords from a profile of interest are compared against the keywords occurring in the content of the specific Web page. Text analysis is also used to screen search terms from search queries.
- **Authorization.** Encryption, password protection, and credit card validation techniques are used to authenticate that a user has the authorization to access given services or data.

---

[2] Balkin, J., B.Noveck and K.Roosevelt. 1999. Filtering the Internet: A best practices model. In *Protecting our children on the Internet: towards a new culture of responsibility.* Bertelsmann Foundation Pub. [www.stiftung.Bertelsmann.de/internetcontent/english/frameset_home.htm]

- **Activity tracing.** Internet usage can be traced by using the server log files and other data logs. These files store details of all Web accesses and can be used to analyze Web-related activities.

The filtering applications that are built on these underlying technologies have been classified as follows:

- **Special purpose browsers for children.** Browser applications can be written that are targeted to child users. Such applications can provide easier search strategies and friendlier graphics, remove advertisements, and provide filtering and search-safe domains in a way that makes it transparent to the user.
- **Child-friendly search engines and portals**. The idea behind both special purpose child-friendly search engines and portals is to use a third party gateway to Web content. Child-friendly portals are Web access sites that try to provide a domain of safe sites for the user to explore. As long as the user comes in through the portal, they view a pre-selected domain set of the Web.
- **Proxy applications.** Proxy software is software that is at the ISP and acts as the intermediary between the client or browser and the Internet. Application software can be added to server proxy modules that permits the execution of text analysis and URL list comparisons for each Internet browser request and response.
- **Activity monitors**. Rather than restrict or control access to Web sites proactively, these applications monitor and log Internet activity for parental review.
- **Restricted access applications**. Applications residing on the host site can be written that restrict access to services or data on that site to authorized users. These applications may encrypt the data so that only authorized users can decrypt and view the data.
- **Non-HTTP applications.** In addition to Web page access, applications can be written using these core technologies to filter content of email and to control access to ftp sites, telnet hosts, discussion and chat groups, and newsgroups.

Potential of the Core Technologies

All of these core technologies have a role to play in filtering mechanisms for the Web. None of these core technologies provides a long term solution on its own. Systems will need to combine the technologies in innovative ways to provide effective solutions. In particular, we note that:

- Site labeling systems are the most flexible and perhaps hold the most promise for the future. Labels may be assigned by the content provider, third party rating services, communities of users, and/or individual users. It is the responsibility of the ISP and/or of the client to use the labels that have been assigned.

- URL lists are the most effective in controlling domains of access. This method is particularly good for creating child-friendly sites. However, the use of lists does not provide the flexibility of labels and the Web is growing so quickly it is very hard to keep lists up-to-date.
- Accurate automated analysis of the content of Web sites is problematic at best, due to the vagaries of natural language, the difficulties of cross-language filtering, and the difficulties of determining the content of graphics and images. Although using this technology to *assist* in the labeling of sites based on text categorization techniques does hold some promise.
- Access authorization can be effective as a reverse filter, restricting who can have access to a given site.
- Activity tracing can only monitor what has been done, it cannot actually filter out any material.

Although Web crawlers are not considered a core technology, they can act as rating agents that rate sites proactively based on content analysis algorithms. Although this has not proven very accurate to-date, there is potential for improvement in this area.

It must be emphasized that none of the above technologies are 100% effective and that the content of the Web is, by its very nature, volatile. In order to be (more) effective, these technologies have to be used in combinations and in layers, both at the ISP and client. As these applications and core technologies are used in combination and as most can be applied at either the ISP or the client or both, there is no clear recommendation as to the best method or where it is best applied.

The areas that hold the greatest potential and where future efforts should be focused include:

- The development of architectures that combine various mechanisms to work in collaboration or as layers within the architecture. For example, labels managed by a Label Bureau, with lists managed by the ISP, and with content analysis at the client.
- Further research into Web crawlers and agents for the proactive categorization of Web sites.
- The development of architectures supporting collaborative filtering for communities of users.

While virtually all of the techniques reviewed in this report can be implemented by an ISP, it is important that the use of filtering is transparent to the user. The user should be informed when filtering is in place or has occurred. To be effective, it is important that:

- The user knows when content has been filtered and why.
- The user knows the criteria for filtering, i.e., what is on the list or in the filter.

Given the dynamic nature of the Web, a concerted and continuing effort into the development, evaluation, and maintenance of filtering and access control mechanisms will be required at all levels, including; government, community, ISP, and individual.

1. Introduction

This project investigates the set of mechanisms that Internet Service Providers (ISPs) have the *option* to provide and that users can *choose* to utilize in order to filter the content delivered to users over the Internet and to allow authorized access to that content. The report is purely descriptive of the mechanisms available and does not provide policy or legal advice or recommendations. It was commissioned by Industry Canada to help promote the development, awareness and use of tools and technologies that enable Internet users to make choices about the content that they access on the Internet.

The Internet is quickly approaching the status of a mass medium in Canada and in so-doing is bringing about some of the same challenges that arose as traditional media, such as television, became a pervasive part of modern life. In particular, the Internet provides a means by which offensive and objectionable content can enter the homes of millions of Canadians. This easy access to a variety of content that some individuals find objectionable is causing significant public concern. In fact, it is increasingly being recognized that the existence of offensive online content – albeit a very small portion of the available online content – may be a significant barrier to the growth and development of the Internet in Canada.

It is also the case that the Internet is not conducive to traditional forms of content controls such as broadcasting regulations. In the first instance, the Internet is not conducive to regulation due to its international nature and underlying packet-based communication mechanism. Any attempts to regulate the flow of content on the Internet at a national level would be immensely expensive, detrimental to the performance of the network and easily circumvented (one can simply obtain Internet access in a jurisdiction that is not regulated), thus making such regulation impractical if not technologically infeasible. These are among the findings of the Industry Canada-commissioned report Regulation of the Internet: A Technological Perspective[3]

An equally, if not more important, point is that regulation is not considered to be an effective way to address the challenge of offensive Internet content from a public policy point of view. This issue was addressed in some detail in the Canadian Radio-television and Telecommunications Commission's (CRTC) Hearings on New Media[4], which outlines the CRTC's approach to Internet services; namely, that they would not regulate these services under the Broadcasting Act. The CRTC points to several market- and

---

[3]Gerry Miller et al. (March 1999). *Regulation of the Internet: A Technological Perspective* [Online]. Available://http://strategis.ic.gc.ca/SSG/it05082e.html. [2000, January]

[4] Submission to the Hearings on New Media, CRTC. 1999, are available at [www.crtc.gc.ca].

technology-based approaches to addressing the control of offensive Internet content that, in their view, hold greater promise in advancing public policy objectives than does regulation.

Included in the selection of mechanisms noted in Regulation of the Internet: A Technological Perspective are desktop filtering technologies that empower users to filter the content that they are able to access on their own computer. Rather than retrieving items from the Web, filtering systems are used to selectively restrict access to materials on the Web. Such systems may be used to "filter out" Web sites or Web pages that contain material that is objectionable to a community of users. For instance, such systems could be used to restrict access by children to sites containing pornographic material.

The premise behind such filtering is that it is better to miss some material than get offensive material, i.e., precision is more important than recall. However, filtering mechanisms are difficult to maintain in an environment in which new material is constantly being added and new Web technologies continue to be developed. To operate effectively, filtering systems must acquire and maintain, over the long term, accurate knowledge not only about the users but also about the documents.

Despite the fact that these technologies are not 100% effective (in the sense that they can be circumvented by computer-adept people, they cannot be relied upon to effectively filter all undesirable content and they may filter out some content unintentionally), they nevertheless provide an important part of the user's content-control tool kit. The Centre for Media Education's recent report: Youth Access to Alcohol and Tobacco Web Marketing: The Filtering and Rating Debate[5], provides a good critique of content filters, reports on a variety of tests, and provides an extensive inventory of screening, filtering and monitoring technologies.

An evolving set of filtering technologies are those that are deployed at the level of the Internet service provider (ISP).[6] These include:

- "Family-friendly" online services that filter content by limiting the domain to which children have access while on-line.
- Manual screening of content according to an ISP's Acceptable Use Policy.
- Access control mechanisms (such as pre-registration or credit card interfaces).
- Encryption and other encoding techniques.

---

[5] Center for Media Education. (October 1999). *Youth Access to Alcohol and Tobacco Web Marketing: The Filtering and Rating Debate* [Online] Available: http://www.cme.org [2000, January].

[6] McRae, Philip et al. (June 1998). *Blocking Content on the Internet: A Technical Perspective* [Online]. Available: http://www.cmis.csiro.au/projects+sectors/blocking.pdf. [2000, January].

Produced for the Australian National Office for the Information Economy by CSIRO Mathematical and Information Sciences (June, 1998). It provides a description of blocking technologies that can be utilised by ISPs.

- Filtering and blocking technologies (including the emerging family of cache-based filtering technologies).

These mechanisms may have significant promise in advancing content management, especially if they place some control in the hands of users. They may also provide a more technologically robust system upon which parents can place a greater reliance on than desktop filters (that may be more prone to tampering).

In developing filtering strategies, one must be aware that there are several categories of participants in filtering, ranging from the content producers to the content consumer.

- Content producers may provide self-referential meta-data or labels that describe the content of their pages that can be used by others to exclude or include such pages.
- Third parties may provide a trusted rating service for users. Generally, such third parties would provide assured labeling for pages, either by reading and rating sites or by accumulating community ratings for sites and providing this information as needed to others. Third parties may provide as a service lists of appropriate Web addresses (URLs), or lists of inappropriate URLs based on a rating of the site content.
- Internet Service Provider (ISP)'s may provide a service to their customers that either blocks access to sites or provides a list of preferred URLs for the users of that ISP. Typically, the ISP's would use rating labels (created in-house or from a third party service), use lists of preferred URL's or blacklisted URL's (created in-house or from a third party service), or use algorithms to filter requests based on the actual content of queries or retrieved pages.
- Clients may exercise filtering at the desktop as well. Typically desktop software would use rating labels and lists to accept or block URL's, or block pages by examining the content of retrieved pages.

Although some filtering mechanisms may be deployed at the desktop, or client side, it is the set of mechanisms that ISPs have the *option* to provide and that users can *choose* to utilize, that is the subject of this investigation.

This report looks at the filtering of Web pages primarily, and e-mail, ftp, telnet, discussion groups, chat groups, newsgroups such as Usenet lists, or instant messaging applications, secondarily. Chapter 2 presents a two-layered classification scheme for filtering technologies. The two layers represent the core technologies that can be used in filtering and the types of applications that use these technologies in filtering. Chapters 3 discusses these layers in detail. Chapter 4 provides an assessment of these technologies and applications while Chapter 5 discusses where the filtering should take place. Chapter 6 summarizes and presents recommendations on the use of these technologies.

2. Classification Scheme for Filtering Technologies

2.1. Two Layer Scheme

A classification scheme to describe filtering technologies consists of two layers: a core technology layer and an application layer. The core technology layer includes categories of core technologies that may be used separately or in combinations by the applications in the application layer. The core technology layer includes: site labels, "black" and "white" URL lists, automatic text analysis, access authority control, and activity tracing. The application layer includes: proxy software, special purpose browsers, special purpose search engines, activity monitors, and access control applications.

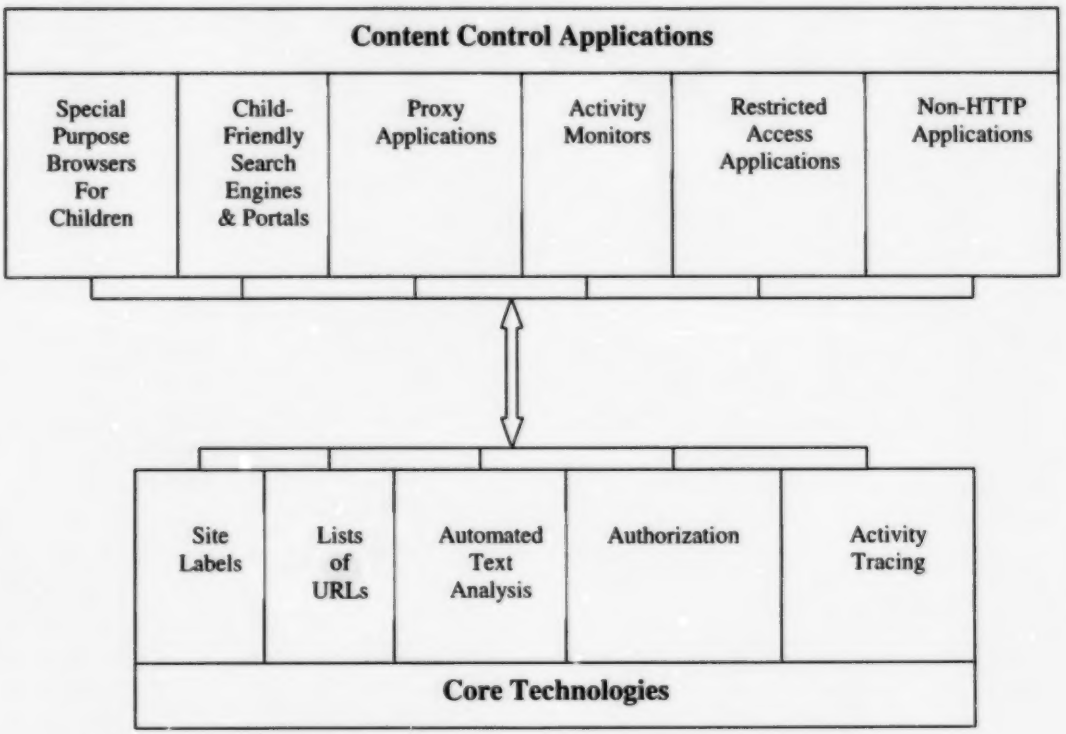

Figure 1. Application and Core Technologies

2.1.1. Core Technologies

The following categories of core technologies will be used in this report to describe techniques for filtering on the Internet today.

- **Site labels.** Labeling refers to schemes to assign content related labels to URL's and/or specific Web pages. Individual rating protocols exist, in general, separate from products or applications using these ratings. In general, these labels can be stored as part of the Web page or separately from the Web page in a database. Labels may be the result of self-rating, third-party authority rating, or community rating by interested users[7].
- **Lists of appropriate or inappropriate sites ("white" and "black" lists).** The most frequently used filtering technology is the use of lists of acceptable and/or unacceptable URL's. "White" lists are used to define a domain of "safe" Web sites within which users can browse. These typically require people to search and select sites that are approved by the provider of the list. "Black" lists are lists compiled of URL's from which requests will not be serviced. The lists are compiled as a service by individuals or by communities of raters.
- **Automated text analysis**. Another way to analyze a Web site is to use software that scans the text of a site to determine the relevance or suitability of pages. Users or groups of users have profiles of interests (positive and/or negative), consisting of keywords and phrases, that are used in this determination. Almost all content based filtering uses some variation of keyword matching, where keywords from a profile of interest are compared against the keywords occurring in the content of the specific Web page. Text analysis is also used to screen search terms from search queries.
- **Authorization.** Encryption, password protection, and credit card validation techniques are used to authenticate that a user has the authorization to access given services or data.
- **Activity tracing.** Internet usage can be traced by using the server log files. These files store details of all Web accesses.

2.1.2. Filtering Applications

Filtering mechanisms are implemented in software applications that may reside at the ISP, client, or some other host. These applications employ one or more of the core technologies and may be described roughly in the following categories.

- **Special purpose browsers for children.** Browser applications can be written that are targeted to child users. Such applications can provide easier search strategies and friendlier graphics, remove advertisements, and provide

[7] J.Balkin, B.Noveck, K.Roosevelt Filtering the Internet: A best practices model. In *Protecting our children on the Internet. Towards a new culture of responsibility.* J.Waltermann and M.Machill (eds). Bertelsmann Foundation Pub. 1999.

filtering and search-safe domains in a way that makes it transparent to the user.

- **Child-friendly search engines and portals**. The idea behind both special purpose child-friendly search engines and portals is to use a third party gateway to Web content. Child-friendly portals are Web access sites that try to provide a domain of safe sites for the user to explore. As long as the user comes in through the portal, they view a pre-selected domain set of the Web.
- **Proxy applications.** Application software can be added to server proxy modules that permits the execution of text analysis and URL list comparisons for each Internet browser request.
- **Activity monitors**. Rather than restrict or control access to Web sites proactively, these applications monitor and log Internet activity for parental review.
- **Restricted access applications**. Applications residing on the host site can be written that restrict access to services or data on that site to authorized users. These applications may encrypt data so that only authorized users can decrypt and view the data.
- **Non-HTTP applications.** In addition to Web page access, applications can be written using these core technologies to filter content of email and to control access to ftp sites, telnet hosts, discussion and chat groups, and newsgroups.

In general, filtering technologies can be applied proactively or reactively. These mechanisms are not disjoint but the application philosophy differs. ISP's and third party authorities provide proactive services to assess the content of websites and accumulate content filtering information. Clients and ISPs generally apply reactive mechanisms to use this content filtering information to encourage or block access to material on the Web.

Proactive or pre-search filtering mechanisms refer to pre-search result activities that collect information that may be used to prevent the retrieval of inappropriate Web pages or to define domains of appropriate Web sites for communities of users. Proactive filtering mechanisms include: the creation, storage, and use of labels, and the creation of lists of good and bad URL's. Both labels and URL lists require significant investment in resource identification and the discovery of both suitable and unsuitable sites.

Reactive or post-search filtering refers to the analysis of URL domain names, metadata, or content after a Web search has been executed. Post search filtering mechanisms may also use URL lists and assessment of label information associated with a given page. In addition, post search may use a content based filtering technique to analyze the actual content of pages returned by the search and on that basis allow or restrict access to pages. The following tables summarize the roles of participants, site owners, third parties, ISPs, and clients, in both proactive and reactive activities. In Table 1 we see that all participants can engage in discovery activities, from self-reporting to dynamic discovery at search time, or combinations. In Table 2 we see that reactive activities are largely handled by software either at the ISP or client, or some combination.

Proactive	**Content Provider**	**Third Party Rating**	**ISP**	**Client**
Site Labels	Self Reporting	Search/create		
URL Lists		Search/create	Create	Create
Text Analysis		Automatic Rating	Web Crawl	
Browsers				Yes
Search Engines			Filter queries and sites	Filter queries And sites

Table 1. Proactive Filtering Mechanisms

Reactive	**Content Provider**	**Third Party Rating**	**ISP**	**Client**
Site Labels			Filter results	Filter results
URL Lists			Filter results	Filter results
Text Analysis			Filter results	Filter results
Browsers				Filter results
Search Engines			Filter results	

Table 2. Reactive Filtering Mechanisms

2.2. Access Modes

Data is available on the Internet in a wide variety of access modes, using both the native Web protocol, HTTP (Hypertext Transport Protocol), and non-HTTP protocols, including: email, ftp, telnet, discussion groups, chat groups, News servers, bulletin boards, and instant messaging services.

Browsers. Browsers, such as Netscape and Internet Explorer, are generally used to access Web pages, located by URL's, using the HTTP protocol. The categories of access control mentioned above all apply to HTTP access. The standard HTTP protocol requires an HTTP header describing the request or response plus a document file.

E-mail - Email may be used by stand alone mail applications or as an application of the browser. Email uses different protocols, usually POP (Post Office Protocol) or SMTP (Simple Mail Transfer Protocol). Most e-mail applications allow the "attachment" of files within the email message. Restriction of access to email follows two main themes: filtering incoming messages from designated sources or by restricted subject matter, and filtering by analysis of actual content of the email message. Attachments, which may be images or videos, for example, are extremely difficult to filter effectively.

News Servers -The Network News Transport Protocol (NNTP) supports News group servers, such as the ubiquitous Usenet, which has been operating since 1979. Usenet news is a bulletin board style broadcast system using the NNTP protocol to let anyone post messages to one of thousands of subjects within a subject hierarchy. NNTP servers then make these messages available to members of the newsgroup. The news files are fed to the news servers continually and so content filtering is possible on a per file basis. Additionally, which newsgroups are brought to the server and which newsgroups users have access to may be subjected to restrictions.

File Transfers - FTP or File Transport Protocol is a TCP/IP protocol that allows Internet users to easily transfer files from one machine to another. TCP/IP (Transmission Control Protocol/Internet Protocol) is the set of protocols for communicating over the Internet, ensuring that messages are routed properly and that those messages arrive intact. The requester does not necessarily have user privileges on the host site. The host site, called the FTP server, may allow general or anonymous access to files or restrict access to its files requiring a userid and password. Most browsers will also initiate a FTP session for URL's beginning with "ftp://".

Telnet - Telnet is another TCP/IP protocol that lets users on one computer log onto accounts on other computers. Unlike HTTP or FTP, the telnet protocol gives the user full access to applications and data on the host computer.

It is relatively difficult to control for FTP and telnet access as these applications run outside the typical browser. Access control lists of acceptable or not acceptable hosts may

be used. As much of the data transferred may not be plain text, filtering of content is problematic at best.

Discussion groups - Discussion Groups are HTTP variations of Newsgroups, where often the membership and/or topic is narrow. For example, a working group or class might set up its own discussion group site. A discussion group lets its members communicate asynchronously posting questions or comments while other members post answers or reflections in an hierarchical manner.

Chat - A Chat Group is a synchronous or real-time variation of discussion groups where comments are exchanged in real-time on some topic of current interest to the active members of the group. Typically, all users currently "in" the chat room may view all of the messages. The resulting exchange may or may not be stored.

Instant messaging - ICQ is the foremost of the current real-time instant messaging applications. ICQ and ICQ-like applications not only act as a chat group allowing real time message exchanges but alert members on who else is currently online.

The access control concerns with Discussion Groups, Chat Groups, and ICQ center on the appropriateness of the community of users of the particular group. Parents may then choose to restrict all access to these applications or to designate appropriate access.

3. Description of Filtering Technologies

In this section we provide a description of the underlying technologies that are used in the various filtering mechanisms. In Section 4.1, we describe the core technologies. In Section 4.2, we describe the application categories.

This report does not examine the implications of social and political concerns, which are critical to the successful deployment of any policies. The success of the technologies rests on the linkage between community values and the development of policies that can be implemented by the technology. The technology should be used to implement community values rather than drive these values.

3.1 Core Technologies

3.1.1 Site Labeling Systems

Labeling refers to schemes to assign content related labels to URL's and/or specific Web pages. The individual rating protocols exist, in general, separate from products or applications using these ratings. In general, these labels can be stored as part of the Web page or separately from the Web page in a database. Labels may be the result of self-rating, third-party authority rating, or community rating by interested users[8]. Self-rating refers to the practice of Web page creators to provide a label describing the content of those pages. The quality of these labels reflects the self-interest of the producer. Third-party authority rating refers to groups or businesses that take on the service of providing labels that describe the content of Web pages for the use of customers or their community of users. Community rating schemes are the more informal efforts by like-minded consumers of Web pages to share their knowledge of the content of Web pages with each other.

The use of labels for content filtering requires several steps: creation of labels, storage of labels, maintenance of labels, dissemination of labels, authentication of label values, and finally the use of available labels by the filtering software to classify given Web pages as acceptable or not acceptable for the given context.

3.1.1.1 Creation of Site Labels

Label Protocols

Although different labeling protocols or schema are available, only a few are in wide use on the Internet.

---

[8] Ibid., p. 10.

The Recreational Software Advisory Council, now part of the Internet Content Rating Association[9] (ICRA) uses a voluntary self-rating schema, the RSACi[10] rating system. The RSACi system rates sites on 4 categories (language, nudity, sex, and violence) and provides users with information on the levels (0 to 4) of these categories in given sites. RSACi is now used by Internet Explorer, CyberPatrol, and CompuServe (in the USA and in Europe).

SafeSurf[11] is an eleven category rating scheme (age range, profanity, heterosexual themes, homosexual themes, nudity, violence, (sex, violence, and profanity), intolerance, glorifying drug use, other adult themes, gambling). In both rating systems, each major category is subdivided into multiple subcategories. Safe Surf is used by both Netscape Navigator and Internet Explorer[12].

Adequate.com[13] is a rating schema based on the USA Television Rating Systems. Sites intended for children are labeled for all children or for older children. Other sites can be labeled for: general audience, parental guidance, parents strongly cautioned, or mature audiences only.

The Entertainment Software Rating Board (ESRBi)[14] has a rating schema for Web sites with games. The categories in this schema include: early childhood, kids to adults (now called everyone), teen, mature, adults only. Each of these categories may be expanded to include content descriptors including: level and type of violence, use of language (mild or strong), type of sexual content, and appropriateness for young children.

The Dublin Core[15] is the most well known schema for resource identification of Web documents. The Dublin Core has a predefined set of descriptors that are used to label the content of Web pages. The current Dublin core set of descriptors are: Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, Rights.

Label Implementation Languages

Once a label schema has been chosen by a service or a community, then a language, such as PICS (Platform for Internet Content Selection[16]) or HTML metadata, is required to

---

[9] *ICRA Internet Content Rating Association*. 2000. [Online]. Available: http:// www.icra.org [2000, January].
[10] *RSAC Recreational Sofware Advisory Council*. 2000. [Online]. Available: http://www.icra.org. [2000, January]
[11] *SafeSurf* [Online]. Available: http://www.safesurf.com. [2000, January]
[12] Center for Media Education. (October 1999). *Youth Access to Alcohol and Tobacco Web Marketing: The Filtering and Rating Debate* [Online] Available: http://www.cme.org [2000, January].
[13] *Adequate* [Online]. Available: http://www.adequate.com. [2000, January].
[14] *ESRB Entertainment Software Rating Board*. 1999. [Online]. Available: http://www.esrb.org [2000, January].
[15] *Dublin Core Metadata Initiative*. 2000. [Online]. Available: http://purl.org/dc [2000, January].
[16] *Platform for Internet Content Selection (PICS)*. 1999. [Online]. Available: http://www.w3c.com/PICS. [2000, January].

implement that schema for individual Web pages. HTML (Hypertext Markup Language) is the set of tags and rules for the use of those tags used in marking up Web documents to be interpreted by the browser for presentation). Often XML can be used to define these languages for use on the Internet, as shown in the figure below. XML (extensible Markup Language) is a language for creating markup languages.

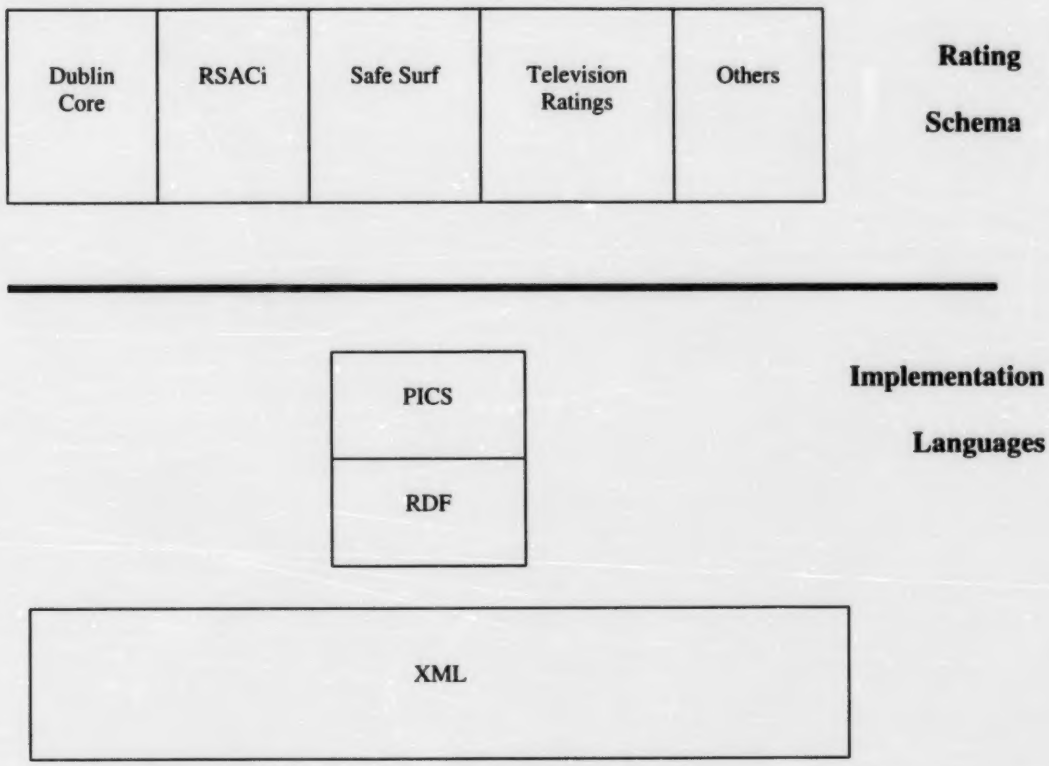

Figure 2. Label Protocols and Implementation Languages

The PICS protocol for labels associated with Internet content was developed in 1995 and continues to be developed by the W3 Consortium. The PICS platform is one on which other rating services and filtering software have been built. It is used by software to label page content for specific uses, such as determining the suitability of individual Web pages for children on the Internet. PICS is not filtering software but can be used by filtering or rating services to develop common categories of labels. PICS is a set of rules that can be used to define labels that describe the content of Web pages and can be shared by a community of users. Shared software can then use the PICS rules to interpret the labeling data for filtering purposes.

The following is an example of a PICS label list, adapted from the W3 example. In this example, the Label Bureau (a separate database of labels) is found at http//www.abc.org/r1.5. Further, a label for the site at *http:www.xyz.com/new.html* was created by John Doe and is valid from Nov 5, 1998 until Nov.1, 2000 and the rating for this site was low (good) for violence and blood but high (poor) for language.

```
(PICS1-1 "http//www.abc.org/r1.5"
        by "John Doe"
        labels on "1998.11.05"
              until "2000.11.01"
        for http://www.xyz.com/new.html
        ratings (violence 2  blood 1  language 4)
)
```

The W3 Consortium has also developed two enhancements for applications that use PICS protocols. The first, called PICSRules[17], lets PICS-compliant applications download third-party rules that can be used by that application to process the PICS data to control access to Web content at that site. The following example of a blocking PICSRule using URL's would block access to any Web page in the domain, http://www.xyz.com.

```
(PicsRule-1.1
        (Policy (RejectByURL (http://www.xyz.com:*/*)
         Policy (AcceptIf "otherwise")
        )
)
```

The following is an example of a blocking PICSRule that would be used by an application. The label information would be downloaded from a third-party Label Bureau.

```
(PicsRule-1.1
        (serviceinfo (
                http://www.lablist.org/ratings/v1.html
                shortname "PTA"
                bureauURL http://www.lablist.org/ratings
                UseEmbedded "N"
                )
        Policy (RejectIf "((PTA.violence >3) or (PTA.language >2))")
        Policy (AcceptIf "otherwise")
        )
)
```

[17] *PicsRules1.1 W3C Recommendation 29 Dec 1997.* 1999. [Online]. Available: http://www.w3c.org/TR/PR-PICSRules.htm. [2000, January].

In this example, labels embedded in documents are ignored and only labels retrieved from the Label Bureau site (http://www.lablist.org/ratings) using the rating scheme "v1.html" are used to assess the documents.

The second PICS enhancement is called Dsig[18] and this allows PICS-compliant applications to verify the authenticity of PICS labels. That is to verify that a statement about a resource, i.e., a label, is true.

Two PICS rating systems that are currently in wide use are those developed by the Recreational Software Advisory Council (RSACi)[19] and by Safe Surf[20], described above.

The RDF (Resource Description Framework[21]), developed by the W3 Consortium, enables machine-understandable descriptions of Internet resources, text, graphics, video, etc. RDF uses the e-Xtended Markup Language (XML) for tagging Web site content, author, target audience, etc. RDF provides a consistent way to provide metadata with Web documents. This information can be used for determining what resources are available, rating content, and classifying documents. Users and groups of users can develop a shared schema that meets their needs using PICS. RDF-enabled applications can then use this metadata for filtering. The Dublin Core descriptors are most often used with RDF.

A sample RDF document descriptor follows:

```
<RDF>
<Description about-"http://cs.dal.ca/index.html">
<DC:title>CS faculty page</DC:title>
<DC:creator>J.Jones</DC:creator>
<DC:date>2000-01-03</DC:date>
<DC:subject>university courses</DC:subject>
<DC:audience>University level students</DC:audience>
</Description>
</RDF>
```

Software[22] can be written to automatically generate metadata in RDF format to describe HTML documents. This RDF metadata can then be used in the filtering process.

HTML META tags may be inserted into Web pages. A Web document is typically HTML tagged text, where the display characteristics of the page are defined by tags, such

---

[18] *DSig 1.0 Signature Label Specification: W3C Proposed Recommendation 03-April-1998.* 1999. [Online] Available: http://www.w3c.org/TR/REC-Dsig-label/. [2000, January].

[19] *RSACi Recreational Sofware Advisory Council.* 2000. [Online]. Available: http://www.icra.org. [2000, January]

[20] *SafeSurf* [Online]. Available: http://www.safesurf.com. [2000, January]

[21] *Resource Description Framework (RDF).* 1999. [Online]. Available: http://www.w3.org//RDF. [2000, January].

[22] Jenkins, C., et al.. Automatic RDF metadata generation for resource discovery. *In Proceedings of WWW8 Conference.* Toronto, May 11-14, 1999. Elsivor Science. 227-42.

as <P> or <B>. An HTML document may have both a head, defined between <HEAD>...</HEAD> tags, and a body section, defined between <BODY>....</BODY> tags. Within the head of an HTML document <META> tags can be used to describe the document content, quality, etc., for the browser and/or server. The data in these tags are not shown on the browser but can be used by the browser to effect the display of the page. Server software may also use the values in these fields as part of the HTTP header fields in a response to requests for this document.

Each <META> tag has two parts, HTTP-EQUIV and NAME, that are used to define a name-value pair, such as "keyword=child". Multiple <META> tags can be concatenated to form a list.

For example, an HTML document may have the following meta-data tags in its HEAD:

```
<HEAD>
        <META HTTP-EQUIV="keywords" CONTENT="nude">
        <META HTTP-EQUIV="keywords" CONTENT="sex">
</HEAD>
```

Upon a request for that document, the server includes the meta data in the HTTP response header and the client software can then use this information to allow or disallow access to this page.

Assignment of Labels

Once a rating scheme has been selected, the next task in the process is the assignment of content labels to Web pages or Web sites. This is the weakest link in the process. The labels may be assigned at a variety of levels, with widely divergent results, including:

- Page creators can self-rate their pages and/or site and provide labels for users of the pages.
- A panel of page raters can examine and rate pages independently of the creators and provide the labels to users as a service.
- Communities of users may individually rate pages and share the label information.
- Computer programs may use text analysis programs to analyze the contents of pages to assign rating labels, typically based on the occurrence of keywords.

3.1.1.2 Managing Site Labels

Once generated and assigned, site labels must be stored in such a manner that they can be accessed. This meta-information may be stored in a separate database to be accessed by queries, it may be encapsulated in a special Web-application recognized format such as the Resource Description Framework (RDF) or it may just be part of the Web page (i.e., as <META> tags).

Storage of Label Information

Once generated, site labels need to be stored in such a way that they can be used by the ISP or client software to perform the filtering tasks. In general, labels can be included in the Web page through <META> tags, associated with the Web page through a mechanism such as RDF, or stored separately in a label database, sometimes called a Label Bureau.

If the labels are included in the Web document then the storage and retrieval of this label information is trivial. It is stored with the document and sent along with the document. Typically, self-describing labels would be stored with the document. If the label information is generated independently from the Web document generation, then it may be stored at the same site or, in fact, anywhere. The Label Bureau then assumes an authentification role for the user of the labels in that bureau.

Control over label information and its retrieval then occurs in the following ways:

- Labels are included in HTML documents by the producer of the document.
- Each ISP maintains a label database as a service to its clients.
- A third party authority provides label information as a service to ISP's or clients.
- Individuals may create and maintain personal label datasets that reside on the client machine and are used as needed.
- Communities of users may contribute and provide feedback to shared label databases held as a service by an organization, ISP, or third party.

Distributing Label Information

The next consideration with respect to site labels is how to distribute the labels so that they are available for filtering as pages are requested.

- The labels can be included in the Web page as HTML META tags. In this case, whenever a page is accessed the labels are available. In the earlier example, client software can easily compare the META-data keywords and exclude any documents with the keyword "nudity". Some browsers, like Internet Explorer (IE), will go to the root document and check for generic labels for documents without specific labels. For example a page with URL …../kidstuff/x1.html may not contain its own META tags, and in this case, the META tags from the parent URL …../kidstuff/ would be used. This label

information can be sent by the server along with Web page as part of the HTTP response header.

- Where labels are stored apart from the Web page they can be used independently. The request for a URL would trigger a database query at the appropriate "Label Bureau" (local, ISP, or third party). A Label Bureau is an HTTP server that may store its own labels or retrieve labels from other sites as needed. The retrieval of the label values and the retrieval of the Web page are separate transactions.
- Similarly, the labels may be stored as RDF files which also must be processed as a separate transaction. RDF files are machine readable descriptions of individual Web resources. RDF files typically use XML to encode the description according to the labeling schema, such as Dublin Core, as described above.
- Labels can be sent by any protocol that uses US-ASCII headers, defined in the Internet mail RFC-822 protocol, such as "subject". Internet email, HTTP, and USENET News are included.
- An extension to the HTTP protocol[23] allows an HTTP client to request that one or more labels be included in the header and sent along with the document requested. Using the HTTP protocol the client can request only a header rather than a header plus a document and so can examine the label information before retrieving the full document.

3.1.1.3 Using Site Labels for Filtering

Finally, filtering mechanisms use the label information to filter access to individual Web pages:

- When the HTML page contains META tags then any of the server, ISP or client may compare the content labels in that page to profiles of interest.
- The HTTP server may include a label header in the response. Some HTTP servers allow an extra header in the HTTP header stream that the client browser can pick up before it processes the data content. In this case, the client application can refuse the data content.
- Labels stored in RDF files may be processed by the ISP or client browser.
- Client software can use local label datasets to either prevent URL requests or to deny access to responses. Labels can be created and stored on client machines to be used by local filtering software.
- Proxy servers may filter requests and responses using PICS rules.
- Label Bureaus at the ISP level may distribute PICS labels for documents on that server. Filtering software at the ISP can then check the Label Bureau for the relevant labels as needed. When the client requests a particular URL the software fetches the document at the same time as it queries the Label Bureau

[23] *PICS Label Distribution Label Syntax and Communication Protocols: Version 1.1 W3C Recommendation 31-October-96. W3C Consortium.* 1999. [Online] Available: http://www.w3c.org/TR/REC-PICS-labels. [2000, January].

database for relevant labels. Depending on the value of the labels for that document it may be blocked or sent.

- Clients or ISPs may subscribe to a rating service that provides labels on request
- A third party may manage labels for users. For example, Adequate.com provides a system whereby Web developers self-rate their content and are provided with a graphic "stamp" that can be included in the Web page and at the same time Adequate.com provides a central Label Bureau for lookup.

3.1.1.4 Examples of Applications that use Site Labels

The filtering software reads the label(s) associated with an URL and allows or denies access to that site based on the rating schema in use. The selection software determines how to use the labels, typically based on actual or presumed user or community preferences.

The use of the PICS structure for encapsulating rating information from a variety of actual rating schema is becoming widely accepted. Currently, the two main browsers, Netscape and Explorer, along with stand-alone services such as CyberPatrol[24] and Adequate.com support the use of PICS.

3.1.2 URL lists

The most frequently used filtering technology is the use of lists of acceptable and/or unacceptable URL's. "White" lists are used to define a domain of "safe" Web sites within which users can browse. These typically require people to search and select sites that are approved by the provider of the list. "Black" lists are lists compiled of URL's from which requests will not be serviced. In almost all cases the lists are compiled by people either working alone or as communities of raters.

3.1.2.1 Creating URL lists

There are several ways that these lists can be compiled:

- Individual users may prepare a local list for their own use.
- Communities of users, school groups, concerned parents, etc., may contribute URL's to a common pool of acceptable or unacceptable URL's.
- ISP's may develop lists based on assumed client preferences that the clients can decide to use to filter content to individual client sites.

---

[24] *Cyber Patrol*. 1999. [Online]. Available: http://www.cyberpatrol.com. [2000, January]

- Third parties may develop lists for communities of users, either on a volunteer or business basis.
- Web crawlers are programs that work autonomously. Given a seed URL, crawlers will request the target document, process that document, and follow all the links embedded in that document. These crawlers can act as rating agents that proactively rate sites based on content analysis algorithms, although this has not proven very accurate to date.

3.1.2.2 Managing URL lists

The lists may be stored at the ISP, on the client machine, or in a trusted third party site for use as needed. The maintenance of up-to-date lists has a direct impact on the usefulness of such a list.

There are certain advantages to maintaining the lists on the ISP. First, ISP may use a special application, called a proxy, through which all client requests to the Internet are processed. Second, updates at the ISP level are shared by all of the clients. Finally, ISPs may maintain multiple lists: general level lists, for example, suitable for children under age 10, community lists in which users have feedback, and customized profile lists for specific customers.

3.1.2.3 Using URL lists for Filtering

Once lists are available their use in filtering is very straight forward. ISP proxy software, portal software, and client application software can all check each request URL and each response URL against the list for each request. Proxy software is software that is at the ISP and acts as the intermediary between the client or browser and the Internet. Portal software is software that also resides at the server site and provides predefined access to certain Internet sites. Client applications, on the other hand, are software that reside on the client desktop, such as the browser and acts on behalf of that client.

A proxy server is a server side application that works with the server software as a channel for client requests. The browser on the client must be configured to use the proxy. The proxy acts as an intermediary in the Web page access process. All client requests go through the proxy server and the resultant content from the Internet goes through the proxy before being sent to the user. The proxy server generally caches or stores the accessed Web pages at the ISP. This cached data may save communication and download time and congestion and can be used for filtering data before it is sent to the user.

3.1.2.4 Examples of Applications that use Lists

SurfWatch[25] – Blocks 100,000 sites deemed inappropriate along with associated IP addresses and virtual domains-representing approximately two million pages; pattern-matching also blocks access to millions of additional pages.

CyberPatrol[26] – uses both black and white lists. The CyberYES list is 3,000 sites selected for their entertainment or education value for children. The CyberNOT list is more than 35,000 sites in 12 categories deemed inappropriate for children. Children cannot access these sites with CyberPatrol installed. In addition CyberPatrol software reads PICS labels and allows filtering to be done solely by PICS labels, if the user chooses.

CyberSitter [27]– maintains lists of categories of "black" lists that are downloaded onto the client machine. CyberSitter filters content deemed as Adult content by all PICS ratings systems. This system employs a range of core technologies along with human rating, including text analysis, for the majority of its blocking.

3.1.3 Automated Text Analysis

Another way to impose a filter on a Web pages is to use software that examines the actual content of the Web pages, as represented by keywords occurring in the text of the Web pages. Users or groups of users have profiles of interests (positive and/or negative), consisting of keywords and phrases, that are used in this determination. Almost all content based filtering uses some variation of keyword matching, where keywords from a profile of interest are compared against the keywords occurring in the content of the specific Web page.

Access to the Web through a browser can be termed as searching or browsing. If searching, the user inputs a query to a search engine which then performs retrieval from previously indexed Web pages. The retrieval is based, normally, on either the Boolean or the vector space model[28]. Both of these models require large inverted files[29] for searching. If browsing, the user clicks on a link in a Web page that has been retrieved previously, either through searching or browsing. The Web page pointed to by the link must be filtered "on-the-fly". Such filtering is done through serial string search algorithms[30] and no inverted files are maintained.

---

[25] *SurfWatch*. 1999. [Online]. Available: http://www.surfwatch.com. [2000, January]
[26] *Cyber Patrol*. 1999. [Online]. Available: http://www.cyberpatrol.com. [2000, January]

[27] *CYBERsitter*. 1999. [Online]. Available: http://www.cybersitter.com. [2000, January]
[28] Baeza-Yates, Ricardo and Berthier Ribeiro-Neto. 1999. *Modern Information Retrieval*. Harlow, England: Addison-Wesley.
[29] An inverted file is an index for each term containing the location of each occurrence of that term in the database.
[30] Frakes, William B. and , Ricardo Baeza-Yates. 1992. *Information Retrieval: Data Structures and Algorithms*. Englewood Cliffs, New Jersey: Prentice Hall.

The Boolean model employs the use of the Boolean operators, AND, OR, and NOT. The query consists of keywords joined together by these operators. For example, if the profile was, "sex OR violence", a Web document would be filtered out if either one or both of these terms occurred in the document. If the profile was "sex AND violence", then both of these terms would have to occur in the document before it was filtered out. The NOT operator could be used in the following manner. If the profile was, "breast NOT cancer", it would filter out those documents that contained the word "breast" and not the word "cancer". If the document contained either both terms or only the term "cancer" it would not be filtered out. Once a set of documents has been identified as satisfying the query, the documents can be ranked on the number of occurrences of the query terms in each document.

The vector space model assumes that the query and each Web document are represented by vectors of keywords. If there are "n" unique keywords in the database and query, then each document and the profile is represented by an n-dimensional vector. The similarity between the query and each document is a measure of the closeness of the vectors within the n-dimensional space. There are a number of similarity measures used but the most common is the cosine similarity measure. This measures the angle between the query and a document. A cosine similarity value of 1 means that they are identical while a value of 0 means the vector representations are orthogonal (completely different). The advantage of this model from the retrieval/filtering perspective is that it ranks the Web documents in the order of their similarity to the query.

String search algorithms require no preprocessing of the Web documents to be searched. There is neither the removal of stop words nor stemming, and no file structures are created. The text of the Web document to be searched is treated as a continuous stream of text and is scanned "on-the-fly". One can think of this model as passing the Web document against the query, as opposed to passing the query against the inverted files of the Web document. These algorithms can find string matches within these documents, even when the string is a substring. For instance, the query term "sex" would be found in the Web document string, "unisex".

Algorithms for the analysis of individual pages may also be used in the automatic determination of rating labels based on the textual content of individual pages. In this case the profile represents the rating category rather than the interests of an individual user or community of users.

The architecture of text analysis systems typically has three modules:

- Profile Manager – maintains profiles of interests from various users or communities of users. For sustainable long-term use, the profile must be updated to reflect changes and refinements in interests and vocabulary.

- Representation of the document - as each document arrives a representation, usually based on keywords, is extracted or generated that then represents the content of that document.
- Selection Algorithm - the document representation is compared against the user profile(s) and the document receives a numeric value representing its similarity to the profile.

3.1.3.1 Creating Profiles and Document Representations

Profiles

Profiles are statements of interest (whether positive or negative) for individual or groups of users. Typically, keywords or phrases are used to represent categories of interest. Weights may be added that indicate the relative importance of each such keyword or phrase. These interests may be long-term or short-term, static or dynamic. Learning algorithms are required to permit the profile to adapt as the content and/or interests shift over time.

Creating the Profile

Critical to the value of a profile is the selection of keywords or phrases that represent the user's interests adequately. The considerations in the selection of those representative keywords include the following:

- The keywords or phrases of the profile should have high predictive power. That is, the keywords should occur very frequently in documents of interest and very seldom in other documents.
- Appropriate stop lists (common words) are used to eliminate words known to have little predictive power, such as "and", "are", or "school" in an educational dataset.
- Stemming algorithms are used to map word variations onto more useful stems. For example, "computing", "computer", and "compute" would map to "comput".
- Both positive and negative words and phrases may be used in the profile. For example, a user might choose the word "arteries" as a positive keyword for a search to do with blood flow. A negative keyword such as "roads" would help filter out documents to do with arteries in the context of public transportation.
- User feedback on the relevance and importance of keywords is essential in the refining and continued management of keywords in a profile.
- Profiles using 50-100 keywords seem to work best as too few leave out important keywords while too many increases the level of noise.
- Special interest thesauri can be used to supplement the user's selection of keywords or expression of interest. Thesauri can also be built for document sets that reflect the term relationships existing in that document set.

- Preliminary profiles of interest are often generated from sample documents selected by the user.

Maintaining Profiles: Learning[31]

The value of profiles in content filtering is directly related to correlation of the vocabulary defining the user interests and the vocabulary used by the producers of the documents. "Learning" techniques are used to refine the profile lists over time to better reflect a user's interests. This may be achieved by using feedback from the user or by analyzing the content of Web pages previously accessed by the user. There are several classes of learning techniques applicable to user profiles:

- Direct learning - the user provides a set of keywords or phrases and periodically updates the set. This puts the responsibility on the user and generally requires a fair amount of effort to be effective.
- Partially direct learning - the system interprets feedback from the user on how well the document presented satisfied the stated interest. The user's effort is increased somewhat during the process but the user does not have an explicit responsibility for the changes in the profile.
- Indirect learning - the system monitors user behavior to rank the relative success of filtering, for example, printing, bookmarking, downloading, or site log usage. This frees the user from explicit intervention but the changes may be ambiguous as the rationale for interest in given documents may not be clear.
- Community filters - a community filter uses the opinions of a community of users to make recommendations. The system maintains a common filter with updates coming from a large number of users of similar interests.

Page Content

Just as the profile consists of a set of keywords representing the interests of the user or the community of users, the content of a Web page is also represented by a set of keywords.

Selecting the Keywords to Represent the Document Content

If string search algorithms are used, the document is viewed as a stream of characters and no stop list or stemming algorithms are applied.

If a model such as the Boolean model or the vector space model is used, keywords are extracted automatically from the text of the document, stop words are eliminated, stemming is performed, and the resulting stems are stored in an inverted file structure. A term weight may be assigned to each remaining keyword stem. The Boolean model assumes a binary weight for the purpose of retrieval. If the term occurs in that document

[31] W.Lam and K.L.Yu. An Intelligent Adaptive Filtering Agent Based on an On-line Learning Model. Proceedings of ACM SIGIR99 Conference. Berkley, Ca. 1999. 287-288.

the weight is 1 and if it does not occur the weight is 0. However, a non-binary weight may be used to determine the ranking of documents retrieved by the Boolean model. If the weight is simply the number of occurrences of the term in the Web document then the weights are static until such time as the document is re-indexed by the search engine.

The vector space model may have either binary weights or non-binary weights. The most robust and well-known metric used to evaluate the suitability of individual words or phrases that occur in a document for use in the document representation is the *Term Frequency-Inverse Document Frequency* (TF-IDF) measure.

TF-IDF is used to evaluate how well the keywords extracted from documents represent the document. The same metric can be used to select words from a sample document for use in the profile. The TF-IDF measure determines the importance of a keyword in representing a document as directly proportional to the number of times the word is used within the document and inversely proportional to the number of documents, indexed by the search engine, in which the word occurs. If a word occurs many times in a document, the word is assumed to be important to that document. If the word occurs in many different documents, then it is probably not a good content descriptor as it cannot be used to discriminate among the different documents.

3.1.3.2 Managing user Profile and Document Representations

As indicated above, the profiles may be static or they may be dynamic. If static, the profiles are constructed once. If dynamic, the profiles must be adaptive, i.e., the profiles must evolve over time to reflect changes in the user's interest and changes in the documents available. In order to be adaptive, there must be some learning algorithm in place.

The document representations are static unless the TF-IDF measure is used. Continual maintenance is required as the calculation of this value for a term is dependent on knowing the number of indexed articles in which the term occurs and the number of documents indexed by the search engine is continually changing.

3.1.3.3 Using Automated Text Analysis to Filter Web Pages

Filtering using automated text analysis techniques can be done in response to a user search query or in response to a user browse request, generated by the user clicking on a link to request a page. In the first instance, the query processing at the search engine is based on either the Boolean model or the vector space model, as described above.

A filtered search can be provided by having a profile that reflects what the user wants to have blocked. In this instance, the profile is treated like a query and retrieval is performed by the search engine. This results in a set of URLs being retrieved that satisfy the profile. The query itself is then processed and a set of URLs is retrieved that satisfy the query. The set of URLs generated in response to the profile is then *deleted* from the

set of URLs returned by the query. The resulting set of URL's is then sent to the user. This can be done either at the search engine or at the ISP acting as the proxy.

However, in a filtering system meant to exclude retrieval of documents, the vector space model is not appropriate for the profile. For rigorous filtering using this model, any Web document having a cosine similarity with the profile greater than 0 would have at least one profile term in it. This reduces to the Boolean model in which the profile terms are all connected by the Boolean OR operator. The Boolean model is generally faster and less expensive to implement and execute.

The user may browse the Web by clicking on a link in a Web page. In this instance, no query is generated. Rather a request is generated to get the indicated page and return it to the user. In this instance, the Web page may not have been processed previously and it is not known if this is an acceptable page or not. Therefore, serial string searching may be done on the retrieved page to ensure that the requested page does not include any of the terms in the profile. In this situation, the filter would have to be enacted at the client or at the ISP, acting as a proxy.

Where Automated Text Analysis Filtering is Applied

These filtering algorithms, although similar in nature, may be executed in a variety of ways:

- They can be executed as part of the software on the ISP proxy server. The software is applied to Web sites which have been retrieved by the proxy server, or are already stored on it.
- They can be applied on a finer grain level on the client's machine as a stand alone application or within the context of a standard browser.
- They may reside on a host Web site.
- They may be executed by one or more agents at the server.
- They may be executed by mobile agents or Web crawlers.

Examples of Applications that use Automated Text Filtering

- Most search engines have an optional filtering mechanism through which the user can request blocking of pages that fall into one or more of several categories, such as sex, violence, etc.
- Surfwatch—automated text analysis filtering is one of the methods employed by this software to find and block objectionable sites from children. Surfwatch also uses URL filtering.
- Cybersitter—uses automated text analysis filtering for the majority of its blocking. It also uses the PICS ratings systems and third party ratings for filtering.
- WebChaperone with iCRT analyzes text for word count ratios, length, structure, and contextual phrase analysis and weighting to detect offensive sites.

3.1.4 Authorization Control Techniques

In addition to filtering search queries and search results, mechanisms are available that require the identification of an authorized user before access to a site or to data will be allowed by the host. This approach complements the filtering out approach by defining domains of material for which explicit access privileges are required. We include a brief description of a sample of these technologies in relation to Internet access.

Encryption – Standard encryption schemes can be used to deliver material to a client site where restricted access to the material is required. For example, adult material that is not suitable for children. Encryption schemes, such as PKI (Public Key Infrastructure) or Private Key Algorithms require the client to be in possession of a valid key in order for the decryption to occur.

Login names and Passwords – Standard login names and passwords may be assigned to authorized users of sites with restricted data. Users wishing to access data or transfer data from these sites must first obtain a login name and password to that site.

Credit card or verification of age – A host site may require a user to provide a valid credit card number or some other proof of age authority before allowing access to the site. The premise is that only persons over the age of 18 would have a valid credit card. Australia[32], for example, has proposed the use of credit card validation to verify that a person requesting access to adult material is 18 years of age or older. Software is now available[33] that is capable of verifying age only without other identifying information.

Biometrics and Hardware – New biometric algorithms and hardware devices are being developed to uniquely identify users. For example, finger scanning, retina scanning, and voice print use a combination of hardware and software to recognize each user. An example of an algorithmic approach is the use of the user's individual typing rhythm[34] of a password to uniquely identify the user.

3.1.5 Activity Tracing

Servers record or log a great deal of Internet use data. Every access through the server is recorded. That log record includes: address of the client, time and date, address of server page accessed, number of bytes, errors generated by the server or by any scripts. These log files, typically used to balance loads and detect usage patterns, can also be used to monitor individual client activities.

---

[32] *Online Services Content Regulation*. [Online]. Available: http://www.aba.gov.au/what/online/restricted.htm. [2000, January].

[33] ZeroKnowledge [Online] Available: http://www.zeroknowledge.com [2000, March]

[34] NetNanny Software International Inc. 1994-1998. *BioPassword: undeniably identified* [Online]. Available: http://www.biopassword.com/docs/BioPassword.PDF [2000, January].

In a similar manner, user activity from keystroke to application command level can be recorded and analyzed to monitor user behavior and screen shots may be taken periodically to record online conversations, email, and word processing activities.

3.2 Application Level

3.2.1 Browser Applications for Children

Browser applications can be written that are targeted to child users. Such applications can provide easier search strategies, friendlier graphics, remove advertisements, and provide filtering and search-safe domains in a way that makes it transparent to the user.

3.2.1.1 Using Browser Applications for Children

The techniques used by these applications include:

- Use of a database of "safe" sites. Children are only allowed access to a database of pre-approved sites. Typically access to these sites is made through a portal interface, icons, or menu lists.
- Use of "black" list URLs or keyword filtering mechanisms. Children are prevented from accessing URLs on the "black" list. These are sites which have been deemed inappropriate for children. The "black" list URLs are not pulled from the Internet. In searching for sites, children are prevented from using keywords which would likely cause inappropriate sites to be returned.

3.2.1.2 Example of Browser Applications for Children

SurfMonkey[35] is an excellent example of a stand-alone browser application designed for children. It blocks using lists and keywords, covering a full range of access types: Web pages, email, chat rooms, and discussion groups.

3.2.2 Child-Friendly Search Engines and Portals

The idea behind both special purpose child-friendly search engines and portals is to use a third party gateway to Web content. Child-friendly portals are Web access sites that try to provide a domain of safe sites for the user to explore. As long as the user comes in through the portal they view a pre-selected domain set of the Web.

3.2.2.1 Using Child-Friendly Search Engines and Portals

Many of the most popular search engines, such as Lycos and Infoseek, provide access modes designed especially for children. These products use one or more of the following techniques:

- Restriction of search keywords.
- Filtering of results.

---

[35] *SurfMonkey*. [Online]. Available: www.surfmonkey.com. [2000, March].

- Directory of "good" sites.

3.2.2.2 Examples of Child-Friendly Search Engines and Portals

InfoSeek — offers links called Disney.com and Kids which are designed to be safe domains for children to explore.
Lycos — also has a link called Kids which children can explore in safety.
Yahooligans[36] — portal /search engine designed for children produced by Yahoo! Inc.

3.2.3 Proxy Applications

A proxy Web server, or proxy, is a software module integrated with the server software that works as an intermediary between the clients and the server (and the Internet). When the proxy is installed, all HTTP access requests are sent to the proxy module before going to the server for processing on the Internet. Without a proxy, all client HTTP requests go to the server directly and are connected to the destination site.

The proxy handles all requests and decides if the request is to be executed. If the requests satisfies the conditions set in the proxy then the request is passed to the server for processing. As individuals can "turn off" the proxy as a user preference option in most browsers, the server may close all outbound connections except to the proxy, so that only the proxy server can make requests over the Internet.

The proxy module typically handles all caching, local storage, of requested documents to improve the efficiency of server operations. The proxy can then, of course, be used to analyze the contents of that material to determine if it satisfies the conditions for delivery to the client.

A Java Applet is a Java program that is intended to be embedded into an HTML document at the server site, transported across a network, and executed at the client site using a Web browser. Proxy modules, such as SurfinGate[37] can be used to evaluate Applets and decide which ones are allowed to go to the client based on policy criteria.

Cache-based Filtering Technologies

Internet HTTP requests from client to server require several steps for each transaction:

- Client makes a connection with the server of interest.
- Client sends a document request message.
- Server responds usually by sending the document to the client. After the document has been received the server breaks the connection.

---

[36] *Yahooligans*. 1994-2000. [Online]. Available: www.yahooligans.com. [2000, January].
[37] Stewart, John. (April 1997). *Proxy Servers to Filter Content*. [Online]. Available: http://webserver.cpg.com/ws/2.3/ [2000, March].

Caching refers to the practice of storing a copy of the document on a local disk or in local memory in order to reduce the latency of subsequent requests to the same page. This means that for the next reference to that document, the document may come directly from a local disk and not necessarily require that the URL server be asked again to fetch and send the same document. This caching can be done at either the ISP server, the client machine, or both. When the ISP uses caching, the cached documents from multiple users can be held in a general cache on the assumption that multiple users may be accessing the same document and, if so, a general cache speeds up requests considerably. At the same time, caching of documents can be exploited for more efficient filtering.

Cache-based applications are designed specifically to handle the caching of Web pages and provide access to the cached pages. They may be used to service HTTP requests from the local cache (if possible), enforce authentication, check labels or other content filtering requirements. Cache-based applications are lighter weight applications than proxy servers and easier to tailor. Although the main selling point of cache-based applications[38] is increased performance on Web page delivery, a significant advantage can also be achieved in increased performance of content filtering.

In this situation, filtering would be done by serial string matching as no inverted file structures are created at the cached site. This would be appropriate for both cases:

- The pages were placed in the cache in response to a search query. Serial string searching of the cached documents as a second layer of filtering on top of the filtering using the Boolean or vector space models.
- The pages were placed in the cache in response to a browser request where the user clicked on a link.

Agents

An agent is a software program that is autonomous, has a goal, and acts on its environment to achieve that goal.[39] The notion of autonomy is important for as the agent has a set of rules that determine if the content of a given site is acceptable or not and hundreds or thousands of agents can be deployed at once to assess the information found at each site and reply back with a simple response. The agents may or may not have any communication between each other and generally operate independently.

Search filter agents[40] generally operate on search results at the ISP proxy level. Search filter agents can be defined for multiple profiles and operate concurrently on the results of a given search. The results of the agents are accumulated by a driving program which uses the individual responses to eliminate unwanted pages.

---

[38] *CacheFlow*. 1998, 1999. [Online]. Available: www.cacheflow.com. [2000, January].
[39] M. Wooldridge. Intelligent Agents. In *Multiagent Systems*, Ed. G. Weiss. MIT Press, Cambridge, USA. 1999. p. 27-78.
[40] Agha, G. and N.Jamali. 1999. *Concurrent Programming in MultiAgent Systems*. MIT Press, Cambridge, Mass.

Filterbots[41] are automated rating agents that evaluate new documents at remote sites as soon as they are published to the Web. The profile of each community of users or individual users are managed by a separate filterbot. The client retrieves documents by communicating with a recommendation server.

Web Crawling (Discovery)

Web crawling is the proactive process of searching out Web resources. That is, to actively seek out pages that are relevant to a pre-defined set of topics, either to include or to exclude. Rather than indexing or using indices to all known Web pages, a Web crawler analyzes its current URL for potential links to other URL's, finding those most likely to lead to relevant pages and excluding the large areas of irrelevant pages. Research[42] has shown that topic-guided crawling is very effective for building high-quality collections of Web documents.

3.2.4 Activity Monitoring Applications

Applications are now available that monitor and log Internet activity for parental review. These applications may be run at the ISP or at the client. ISP's can easily log all user activities through the server. At the desktop, applications are available that record particular user activities and prepare views of that activity for analysis. These applications employ some or all of the following techniques:

- Record use of certain keywords or phrases in searches.
- Log all URLs visited.
- Limit the amount of time spent browsing.
- Periodically record a screen shot.
- Capture cached graphics.
- Cookies for saving data on client activity.

For example:

NetNanny—monitors and logs all sites that are visited, programs that are used and words and phrases that are typed and/or received if the user chooses to.
CYBERsitter—will maintain a complete history of all Internet activity including attempts to access blocked material.
Bess—also has "reporting software" to monitor Internet activity.
Smart Alex ICU – monitors all activities, both on and off the Internet using screen shots.

---

[41] Sarwar, B., J.Konstan, A.Borchers, J.Herlocker, B.Miller, and J.Riedl. 1998. Using filtering agents to improve prediction quality in GroupLens Research Collaborative Filtering System. Proc. of CSCW 98 Seattle Washington. 345-354

[42] S.Chakrabarti, M.vanden Berg, and B.Dom. 1999. Focused crawling: a new approach to topic- specific Web resource discovery. In *Proceedings of WWW8 Conference*. Elsivor Science Press. P. 545-562.

3.2.5 Restricted Access Applications

Applications residing on the host site can be written that restrict access to services or data on that site to authorized users. Data may be encrypted so that only authorized users can decrypt and view the data. Please see Section 3.1.4 for more details.

3.2.6 Non-HTTP Applications

Although ftp and email and other non-HTTP data can be accessed through the browser, the potential for these non-HTTP protocol applications to be used outside of the standard browser-server venue to gain access to data is significant. Applications have been built specifically for filtering these types of data. These applications may be stand-alone or incorporated as a feature in larger more general filtering application software.

E-mail - Email filtering applications focus on email spam and email content filtering. Spam is a flood of copies of email messages, often advertising, sent to as many email accounts as possible. Spam e-mail lists are often generated by scanning Usenet postings, searching Web addresses, and gaining access to email lists for other purposes. Spam e-mail lists may include not only individual email addresses but also public or private lists such as those used for discussion groups. Spam filtering uses rules to scan the headers of incoming messages to eliminate likely spam messages. For example, the domain of the sender, the contents of the From or Subject fields are often used.

Content filtering is a finer grained inspection of the message contents to determine if the email is to be accepted for delivery. Most content filtering uses a profile of keywords that is compared against the content of the message, in a manner similar to content filtering of Web pages. Rejected or suspect email may be rerouted to a folder with password access protection.

Email may be monitored for offensive material in outgoing mail as well, especially when associated with a proxy server. Theses applications range from simple filtering of messages based on keyword to more intrusive control over content. For more control, email applications, such as *Email for Kids*[43] give the parent control over both incoming and outgoing messages to monitor and regulate both content and to whom/ from whom the children communicate.

NewsServers – News server filtering, such as for UseNet, is largely accomplished using two core technologies: lists of acceptable/not-acceptable newsgroups to block at the newsgroup level, and keyword based filtering.

FTP – File transfers can only be reasonably stopped either by preventing any ftp access or by controlling the sites to which access is allowed. Individual sites may control to

[43] *Email for Kids*. [Online]. Available: http://www.email-connection.com. [2000, January].

whom access is granted by the use of any of the access control technologies, such as user id and password.

Telnet – Since telnet is remote access of a user to his or her own account on a remote computer the main control mechanism available is disallowing telnet access through a firewall.

Discussion groups, chat rooms, and instant messaging applications – Controlling access to these applications or to lists of acceptable discussion groups and chat rooms is straightforward. ICQ-like applications may be written to adhere to a block list or list of acceptable participants.

Often these control features are bundled with other control features in broader applications such as NetNanny or SurfMonkey.

4. Assessment of Types of Filtering Technologies

4.1 Core Technologies

4.1.1 Site Labels

Advantages

- A rating scheme with layers of ratings may allow more sophisticated use of the label. For example, filtering decisions made by age or level of violence.
- Users can rely on a trusted authority to categorize content fairly.
- Rating schemes can be devised or customized to meet special needs.
- If the rating schemes are open then multiple schemes can be used in conjunction. This means that an application could send a label request to more than one Label Bureau, perhaps one national and one international.
- The use of rating schemes allows both flexibility and consistency in filtering applications.
- Security and Authentication - Labels can include a cryptographic or digital signature so that the software or user can verify the authenticity of the label.

Disadvantages

- It is currently hard to analyze video image and audio content automatically.
- Much of the effective labeling relies on human categorization.
- Labels are inherently biased. They will always reflect the ideology of the group which created the labels.
- Covering significant portions of the Web is costly and slow.
- Maintaining up-to-date coverage is nearly impossible, particularly for sites with changing content. Once a site has been labeled the content provider may change the content without changing the label.
- Labeling virtual documents is impossible. These are pages whose content is generated on the fly.
- Questionable value of Web sites labeled by the Web site producers.
- Granularity of rating scheme. Too many categories leads to difficulties in assigning ratings consistently. Too few categories leads to undifferentiated ratings. For example, "bad language" does not differentiate between racial epithets and sexually explicit language.
- Using a Label Bureau or third party slows responses. Techniques can be used to minimize this, including keeping local copies of labels and making requests in parallel.
- Using embedded HTML labels requires trust of the producer of the label.
- The separation of label and document content may fail under several scenarios. For example, moving a document from one URL to another would "unlabel" that document.

4.1.2 URL lists

"Black" Lists

Advantages

- Easy to implement.

Disadvantages

- Quality of the list and completeness of the list.
- URL's are easily changed.
- Mirror sites may mask actual URL.
- Maintenance of lists is a huge task.

"White" Lists

Advantages

- Easy to implement.
- Robust protection.
- Provide "safe" domains.

Disadvantages

- Quality of the list.
- Exclude many potentially useful sites.

4.1.3 Automated Text Analysis

Advantages

- Web sites may be labeled quickly and easily.
- Keywords and phrases to be searched for may be chosen to reflect a user's personal preferences.
- This technique may be used to analyze text generated "on-the-fly".
- All the Web pages accessed by the user are analyzed.

Disadvantages

- Generates poor results: The context of the word or phrase is not taken into account. For example, if the keyword "breast" is used to filter out pornographic sites, sites about breast cancer will also be filtered out.
- Images, videos and other forms of non-text data are very difficult to analyze.
- The user lists of keywords and phrases must be updated to reflect changes in the users interest and changes in terminology in the Web content.

4.1.4 Authorization Techniques

Encryption

Advantages

- Allows users of a shared facility to accept material that may not be suitable for other users of that facility.
- Encryption systems are readily available

Disadvantages

- Material may be left on the user's disk in unencrypted form.
- Users need to retain and protect their key.

Login Names and Passwords

Advantages

- Site host has the responsibility for providing access authority to its material.
- Passwords are a proven and known technology.

Disadvantages

- Typically users choose passwords that are easily determined by others, particularly people who know them, such as pet names, address variations, birthdays, etc.
- When people choose a less obvious password, they often write it down and leave it near the computer.
- To be effective, passwords need to be changed frequently.
- People with multiple passwords often use only one or two different ones.

Credit Card Age Restriction

Advantages

- No set up required from the user or administration point of view.
- May allow for privacy of access.

Disadvantages

- It is very easy for users under the age of 18 to get a valid credit card number.

Biometrics and Hardware

Advantages

- Allows for unique identification of individual users.
- Removes the need for frequently changing the password.

Disadvantages:

- Hardware based solutions are still expensive.

4.1.5. Activity Tracing

Advantages

- Easy to collect the information.

Disadvantages

- May be difficult to identify the actual user.

4.2 Application technologies

4.2.1 Special Purpose Browsers for Children

Advantages

- Can have child-friendly interfaces.
- Limited experience or skills needed to use the system.
- User cannot "get-around" the controls.
- User is only aware of the domain and activities provided by the browser.

Disadvantages

- Knowledge on how to use this browser may not scale up when using other more standard browsers.
- May over-restrict some children.

4.2.2. Child Friendly Search Engines and Portals

Advantages

- Predefines a "safe" domain of knowledge.
- Combines most useful sites and search engines into a consistent and "one-click" guide.

Disadvantages

- Keyword filtering is not very effective. For example, the first and second try on two popular child-friendly search engines produced graphic sexual material.
- Links from retrieved sites, particularly from portal interfaces, may not be screened.

4.2.3. Proxy Applications

Advantages

- Can control all Internet accesses through server.

Disadvantages

- Relatively easy for users to bypass.

4.2.4. Access Authority Applications

Advantages

- Allows users of a shared facility to accept material that may not be suitable for other users of that facility.
- Site host has the responsibility for providing access authority to its material.
- May be able to use an identification system, such as credit card, to validate age rather than a registration process recording personal identification information.

Disadvantages

- Material may be left on the user's disk in unencrypted form.
- Users need to retain and protect their key.
- Hardware solutions are currently very expensive.
- Typically users choose passwords that are easily determined by others.
- To be effective, passwords need to be changed frequently.
- It is very easy for users under the age of 18 to get a valid credit card number.

4.2.5. Activity Monitors

Advantages

- Provides a history of activity that parents can use to ensure that children are not viewing inappropriate materials.

Disadvantages

- Heavy time investment in identifying misuse.
- After the barn door has closed.
- May violate rights to privacy of users.

4.2.6. Filtering for non-HTTP Applications

Advantages

- Much of the communication and transfer of files is accomplished using these applications.

Disadvantages

- Filtering or blocking for these applications typically requires separate applications.
- Browser users can often circumvent server-based controls by using non-http access applications.

5. Where the Filtering is done

Filtering activities occur either before an actual search is made or after the search has been completed but before any material is presented to the user. From a technological perspective, this processing can take place at the ISP or on individual clients. The decision on where to host the software needs to be based on other factors: individual control, community view vs. personal view, responsibility for generation of lists, keywords or labels, responsibility for updates, and risk analysis. In this section the advantages of hosting the filtering mechanisms on the individual client and at the ISP are discussed. Of course, this is not an either or situation and combinations of mechanisms on clients coordinating with mechanisms on ISPs may offer the best of both as well as increased flexibility.

5.1 Stand Alone Mechanisms (client machine)

At the client level several mechanisms are available:

- Activity monitoring.
- Special browser application.
- Pre-search filter of keywords or phrases.
- Cache-based filtering based on URL, label, or content.

Client Filtering of cached documents[44]

Client applications can examine the content of data retrieved from the Internet after it has been downloaded into the client cache and before it is displayed by the browser. The filtering may be based on specific lists of URL's stored in the client machine, examination of keywords of the documents, or by examining the labeling data of the individual pages. A variation is to download the material and at the same time request labels from a third party associated with those pages for filtering at the client level.

Advantage

- Fine grained personalized and confidential profiles can be used.

Disadvantages

- Objectionable material is already on the client machine and could easily be accessed by other applications.
- Graphics in particular are very difficult to block based on text algorithms.

---

[44] McRae, Philip et al. (June 1998). *Blocking Content on the Internet: A Technical Perspective* [Online]. Available: http://www.cmis.csiro.au/projects+sectors/blocking.pdf. [2000, January].

5.2 ISP Mechanisms

ISP's may offer value-added services to their users by providing filters at the ISP level, using firewalls or proxy servers. Firewalls are gateway machines that control all traffic to or from the host system and the Internet. Software at the firewall can be configured to reject classes of access for all users of the system.

A proxy server is a server side application that works with the server software as a channel for client requests. The browser on the client must be configured to use the proxy. The proxy software acts as an intermediary in the Web page access process. All client requests go through the proxy server and the resultant content from the Internet goes through the proxy before being sent to the user. The control mechanisms, typically accomplished using a proxy server, include:

- Application of label filters.
- Filtering of search results by lists or content.
- Providing a database of "safe" sites for searching.
- Providing portals sites for children.

The proxy server generally caches or stores the accessed Web pages at the ISP. This cached data may save communication and download time and congestion and can be used for filtering data before it is sent to the user. The proxy at the ISP can perform several filtering related tasks:

- Cache certain page hierarchies using lists.
- Check with the source site to verify if the cached page is current before sending to the client.
- Check the labels associated with pages in cache.
- Refuse to send certain pages from cache to the client based on lists.
- Analyze the content of pages to decide if to send or remove.

Most ISPs provide browser software pre-configured to go to an ISP portal site as the users default home page when the browser is launched. In any event, the client may configure his or her browser to "go through" the ISP proxy server. The use of the proxy prevents the clients from going directly to the Internet but rather all requests are processed through the proxy. Proxy software can store or cache results that can then be filtered before results are sent to the client.

Advantages

- Eliminates the need to install separate software at the client.
- Parallel and computationally intense applications may be used at the ISP.
- Filtering at the ISP cache level may reduce the amount of processing necessary.
- ISP may maintain a large database of labels and/or community lists.
- ISP takes on the responsibility of maintaining the lists.

- ISP may deploy proactive search strategies to catalog new sites.
- ISP may filter in a way that is transparent to the client.

Weaknesses

- Sites can easily be renamed and hence not match names in a proxy list.
- Translation services may produce versions that contain words that do not occur in the profile and are therefore not filtered effectively by keyword algorithms.
- Lists must include both domain name and Internet IP address.
- Not all applications work well through a proxy and it may effect performance of such applications.
- Push technologies generally bypass proxy request checks and deliver content directly to user sites.
- Not everyone accesses the Internet through an ISP.
- Mirror sites, alternative URL's for some sites, need to be included in any lists as the same material may have multiple URL's.
- As traffic on the Internet increases, proxies become bottlenecks and may degrade performance.
- If links on external Web pages do not use the standard port number then requests for these pages will not be caught on the lists.

6. Summary and Recommendations

In summary, there are a number of ways that access to content can be controlled, and these rely on one or more underlying core technologies. As these applications and core technologies can be used in combination and as most can be applied at either the ISP or the client or both, there is no clear recommendation as to the best method or where it is best applied. Nonetheless, the users must be informed and involved in this process.

The following is a review of the potential of the core technologies:

- Site labeling systems are the most flexible and perhaps hold the most promise for the future. It is the responsibility of the ISP and/or of the client to use the labels that have been assigned.
- URL lists are the most effective in controlling the domain of access. This method is particularly good for creating child-friendly sites. However, the use of lists does not provide the flexibility of labels and the Web is growing so quickly it is very hard to keep lists up-to-date.
- Accurate automated analysis of site content is problematic at best, due to the vagaries of natural language, the difficulties of cross-language filtering, and the difficulties of determining the content of graphics and images. Although using this technology to *assist* in the labeling of sites based on text categorization techniques does hold some promise.
- Access authorization can be effective as a reverse filter, restricting who can have access to a given site.
- Activity tracing can only monitor what has been done, it cannot actually filter out any material.

Although Web crawlers are not considered a core technology, they can act as rating agents that proactively rate sites based on content analysis algorithms. Although this has not proven very accurate to-date, there is potential for improvement in this area.

These filtering mechanisms may be implemented at various places in the overall architecture of the Web:

- They can be implemented as part of the software on the ISP proxy server. The software is applied to Web sites which have been retrieved by the proxy server, or are already stored on it.
- They can be applied on a finer grain level on the client's machine as a stand alone application or within the context of a standard browser.
- They may reside on a host Web site.
- They may be executed by one or more agents at the server.
- They may be executed by mobile agents or Web crawlers.

The economic cost/benefit analysis of these techniques is beyond the scope of this report. However, such an analysis would examine:

- The cost of proactive categorization or raters of Web sites on an ongoing basis.
- The cost of storage and distribution of this information.
- The cost of processing this information at search time.
- The cost of development of stand-alone browsers incorporating appropriate features.
- The cost of verification of Web sites, labels, lists and users.

It must be emphasized that no filtering technologies are 100% effective and that the content of the Web is, by its very nature, volatile. In order to be (more) effective, these technologies have to be used in combination and in layers, both at the ISP and client.

The areas that hold the greatest potential and where future efforts should be focused include:

- The development of architectures that combine various mechanism to work in collaboration or as layers within the architecture. For example, labels managed by a Label Bureau, with lists managed by the ISP, with content analysis at the client.
- Further research into Web crawlers and agents for the proactive categorization of Web sites.
- The development of architectures supporting collaborative filtering for communities of users.

While virtually all of the techniques reviewed in this report can be implemented by an ISP, it is important that the use of filtering is transparent to the user. The user should be informed when filtering is in place or has occurred. It is important that:

- The user knows when content has been filtered and why.
- The user knows what is on the list or in the filter.

Given the dynamic nature of the Web, a concerted and continuing effort into the development, evaluation, and maintenance of filtering and access control mechanisms will be required at all levels, including; government, community, ISP, and individual.

Appendix

Examples of Filtering Software

The following list of filtering software systems[45] are examples using a range of four control mechanisms.

- Safe Domain—recommend appropriate content for children.
- Search—select content that is appropriate for children and matches a query.
- Monitor—record for later inspection a list of the content accessed or attempted to be accessed by a user.
- Block—prevent children from accessing content.

Notice that by far the most frequently used core technology is blocking using labels, URL lists, or text analysis, and that most systems employ more than one technology.

[45] Cranor, Lorrie Faith, Danielle Gallo and Paul Resnick. (1998). *Technology inventory: a catalog of tools that support parents; ability to choose online content appropriate for their children.* [Online]. Available: http://www.research.att.com/projects/tech4kids/index.html. [2000, January].

No.	Software	Safe Domain	Search	Monitor	Block
1	Access Management Engine (AME)	X			X
2	AltaVista Filtered Search Service		X		
3	America Online Parental Controls				X
4	Bess			X	X
5	Bonus.com the SuperSite for Kids	X			X
6	ChoiceView			X	X
7	Click and Brose Jr.	X			X
8	Cyber Patrol	X			X
9	CYBERsitter			X	X
10	Cyber Snoop	X		X	X
11	Disney's Blast Online	X			X
12	EdView	X	X		X
13	ESRB				X
14	evaluWEB				X
15	GuardiaNet			X	X
16	I-Gear			X	X
17	IBM Web Traffic Express				X
18	The Internet Filter			X	X
19	The Internet Kids & Family Yellow Pages	X			
20	KidDesk Internet Safe	X			X
21	Kids CyberHighway	X			X
22	Mail Gear			X	X
23	Mayberry USA			X	X
24	Microsoft Internet Explorer				X
25	NetFilter				X
26	Net Nanny			X	X
27	Net Rated			X	X
28	Netscape 4.5				X
29	Net Shepherd		X		X
30	PlanetView				X
31	RSACi	X	X	X	X
32	SafeSurf	X	X	X	X
33	Scholastic Network	X			
34	SmartFilter			X	X
35	SurfWatch				X
36	WatchGuard SchoolMate				X
37	WebChaperone with iCRT			X	X
38	WebDoubler			X	X
39	WebSENSE			X	X
40	WinGuardian			X	
41	X-STOP				X

Table 3. List of Filtering Software Systems and their Control Mechanisms